Today I will be showing you some standard and some new visualizations for interpretability tools in the form of static Jupyter Notebooks. I will share my screen with the visualizations and then give you control so you can click around the static notebooks however you like. I will be sharing 2 different notebooks, with visualizations from a commonly used, open source data set. As you are looking at the visualizations, I will be asking you questions to elicit feedback on the standard and new visualizations. I will be recording audio and video of this interview to capture the screen sharing component, but feel free to turn off your camera if you do not want your face captured in the recording. Do you have any questions before we begin?

In each notebook, there is a description of the dataset if you would like to read more. Imagine each notebook was created by a colleague and you are reviewing it. For each notebook, you have 3 objectives: to get a general understanding of what the model is doing, to explain the model to someone who is not a data scientist, and to figure out next steps for debugging. Spend a few minutes exploring the notebook and visualizations, thinking aloud as you look at the visualizations. Then I will ask you some more specific questions about the model and you can refer back to the notebook to answer those questions.

This notebook contains a trained model and explanations for the [Adult/NHANES] dataset with some [standard/new] visualizations. Please think aloud as you are looking at the notebook.

1. Based on what you can see in these visualizations, what do you think the model is learning overall?
    1. What did you use to answer this question?
2. Based on what you can see in these visualizations, how would you explain what this model is learning to someone who isn't a data scientist?
    1. Which visualizations, if any, would you show to someone who isn't a data scientist to help explain?
    2. What did you use to answer this question?
3. If you were to go about debugging this model, what would your next steps be?
    1. What did you use to answer this question?

[Repeat above with new visualizations notebook]

1. What did you like about each new ranking metric?
2. What did you dislike about each new ranking metric?
3. In your work or domain, would you use any of the new ranking metrics?
    a. Which ones? Prompt for each different ranking metric
    b. For what? Prompt with different tasks: communication, collaboration, debugging, determining credibility of the model
4. Was there any additional value or information gained from the new ranking metrics?
5. Do you have suggestions for improvements to the rankings shown?
6. Do you have any suggestions for other rankings?
7. Do you have any suggestions for other ways of visualizing or communicating the global impact of features on the model beyond ranking metrics?